
THISL progress report - 1999sep

Dan Ellis

International Computer Science Institute, Berkeley CA

<dpwe@icsi.berkeley.edu>

Outline

- 1 Spoken queries recognized**
- 2 TREC modspec processed**
- 3 ThisIGUI updated**
- 4 FYI: General audio retrieval**



1

Spoken query recognition

- 23 TREC query sentences (351 words)
- Pronounced by 9 speakers (from BBC) + 1
- Recognized by RNN and RNN+MSG

Speaker	RNN WER	RNN+MSG WER
nch	16.0%	15.1%
mrm	21.4%	21.9%
cdm	24.5%	21.1%
jp	24.5%	24.8%
mac	25.9%	23.9%
mpj	26.5%	24.2%
ljc	26.5%	25.6%
cmb	35.0%	29.3%
cp	38.7%	34.2%
dpwe	36.2%	37.9%
Avg	27.5%	25.8%

- Pass to IR for evaluation ...



2

TREC modspec processing

- **Modulation-filtered spectrogram (msg) features**
 - helped in BN'98 eval
 - could help for TREC8 data?→run msg neural net for probability-combination with RNN
- **Automatic wav→msg→Ina loop**
 - ftp between Sheffield and ICSI
 - 2.5M weight network ran in 0.3xRT on TetraSPERT
 - 900+ files (550 hours) processed in 4 weeks



3

ThisIRGUI updated

Date filters

Speech input

BBC data

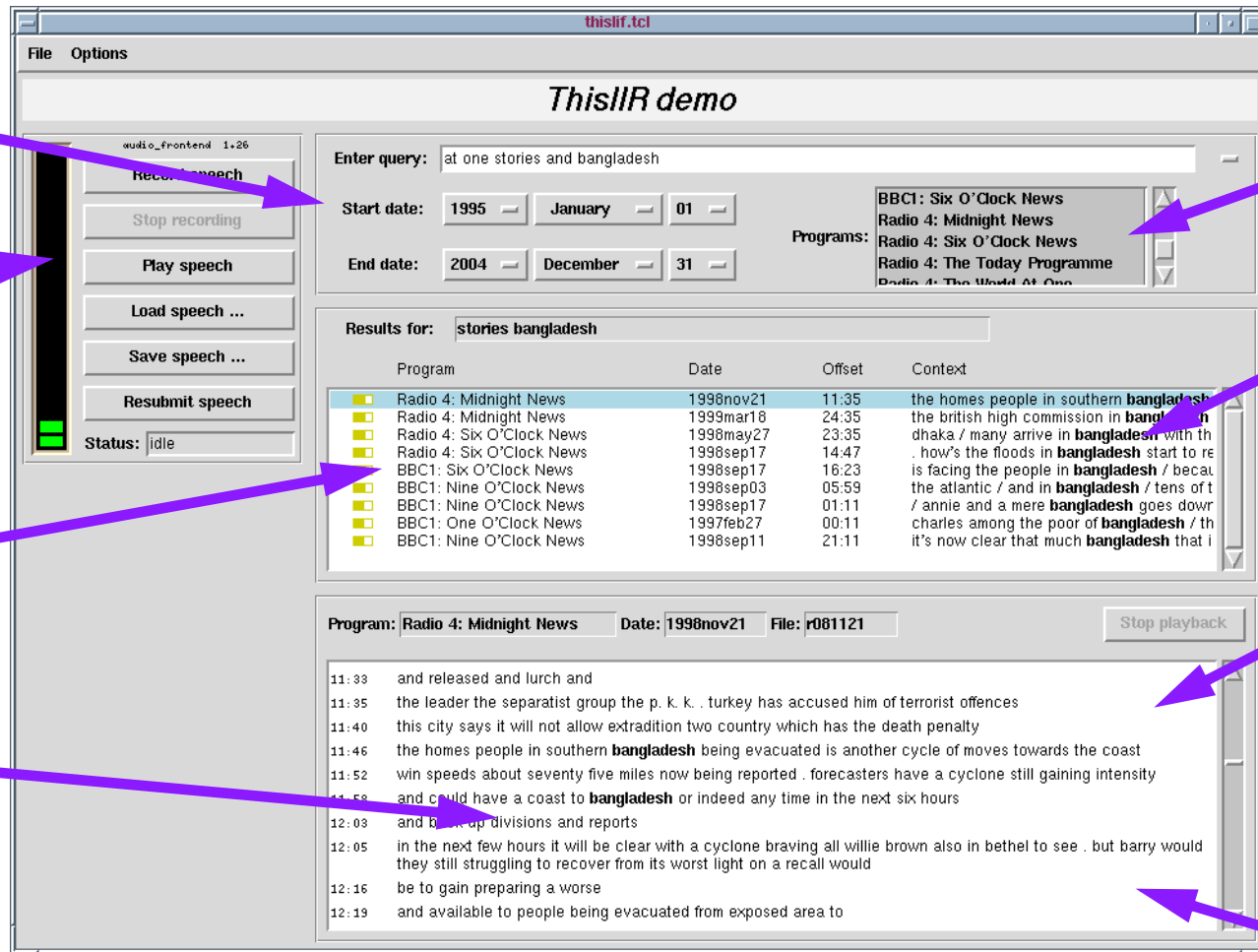
click-to-play

Program filter

thisIR-0.2 & thisIR-0.5 compatible

Pauses & sentence breaks shown

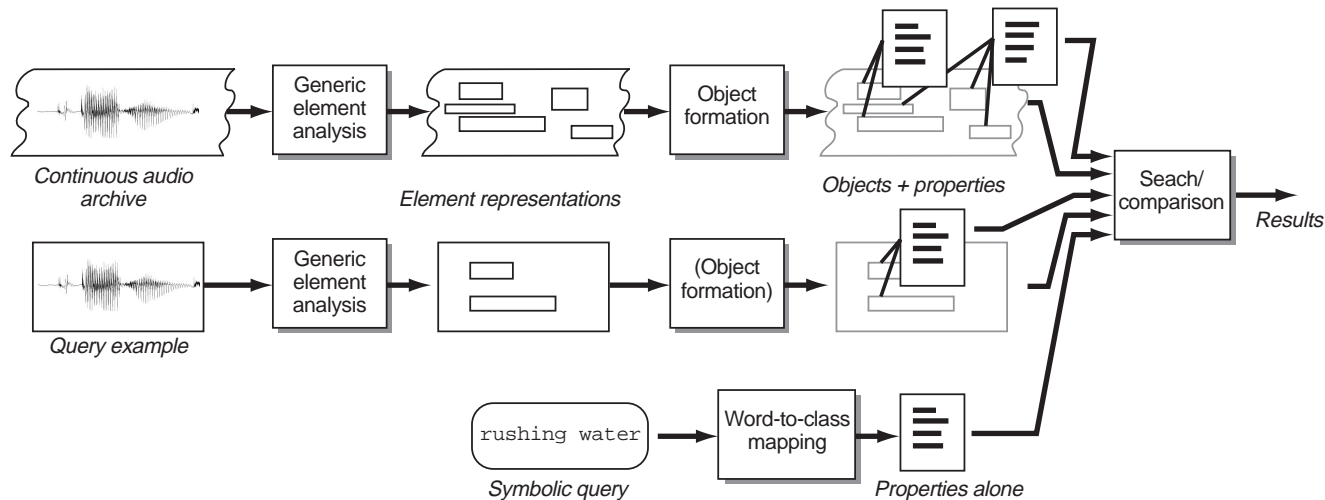
Whole-show transcripts



4

General audio mixture retrieval

- **Information retrieval for nonspeech audio**
 - non-word events (laughter, shouting)
 - sound effects on soundtracks (engines, cars)
 - ambiences
- **Approach: auditory elements, learned patterns:**



- **Just-submitted proposal to NSF**